On the application of cluster analysis for vegetation pollution assessment in the area of mining enterprise

Cite as: AIP Conference Proceedings **2172**, 080006 (2019); https://doi.org/10.1063/1.5133564
Published Online: 13 November 2019

Maxim A. Medvedev, Alexander N. Medvedev, and Dzmitry Melnichenka

On the Application of Cluster Analysis for Vegetation Pollution Assessment in the Area of Mining Enterprise

Maxim A. Medvedev[1,2,a)], Alexander N. Medvedev[1,2,b)] and Dzmitry Melnichenka[3)]

[1]*Ural Federal University, Mira st., 19, Yekaterinburg, Russia*
[2]*Institute of Industrial Ecology UB RAS, Sofia Kovalevskaya st., 20, Yekaterinburg, Russia*
[3]*Belarusian State University of Informatics and Radioelecktronics, Republic of Belarus*

[a)] Corresponding author: medvedevmaa@gmail.com
[b)] medvedev@ecko.uran.ru

Abstract: The assessment of vegetation contamination in the influence area of mining enterprises is an important part of the research during the environment monitoring. There are different statistical methods that can be used for the analysis of data obtained in environmental monitoring. The article presents the results of cluster analysis of the chemical composition of agricultural vegetation samples collected in the area of copper-pyrite ore deposit location. During the analysis, all samples were divided into three clusters. One can suggest that this separation may be due to different mechanisms of pollutants entry into the particular sampling sites, as well as to the location of these sampling sites relatively to the enterprise industrial area. According to the results of the study, it can be concluded that cluster analysis is an effective tool for distinguishing the zones being characterized by different pollution mechanisms of grassy vegetation, when there are a small number of measurements and relatively low levels of the samples pollution.

Keywords: cluster analysis, vegetation pollution, environmental monitoring.

INTRODUCTION

Assessment of vegetation pollution in the area of mining enterprises is an important part of research during environmental monitoring [1-3]. The need for monitoring is provided in the environmental legislation of the Russian Federation and other countries [4].

As part of the implementation of the environmental monitoring program in the Safyanovskoye deposit (Middle Urals, Russia), the Institute of Industrial Ecology (Ural Branch of the Russian Academy of Sciences, Yekaterinburg) conducts a long-term observation cycle to assess the pollution of agricultural vegetation with metals that are specific to the ore mined and location of the deposit [5].

Pollution of vegetation in the area of work is considered as one of the important parameters when conducting environmental monitoring and environmental impact assessment. Assessment of vegetation pollution allows determining the scale of pollution over time and assessing the long-term influence of a mining enterprise on its surrounding territory [6, 7].

Various statistical methods can be used to analyze experimental data obtained in environmental research as well as in social and economic sciences [8-15]. This article presents the results of cluster analysis of the chemical composition of samples of agricultural vegetation, collected in the vicinity of the copper-pyrite ore deposit.

Proceedings of the 45th International Conference on Application of Mathematics in Engineering and Economics (AMEE'19)
AIP Conf. Proc. 2172, 080006-1–080006-6; https://doi.org/10.1063/1.5133564
Published by AIP Publishing. 978-0-7354-1919-3/$30.00

METHODOLOGY AND RESULTS

Sampling and research of environmental pollution (including vegetation) is carried out annually as part of local environmental monitoring of the state and pollution of various environmental components in the development area of the Safyanovskoye copper-pyrite deposit (Figure 1). At the moment there is a significant amount of accumulated monitoring data that are aggregated in the information system of environmental monitoring [16] and subsequently analyzed together with the newly received data.

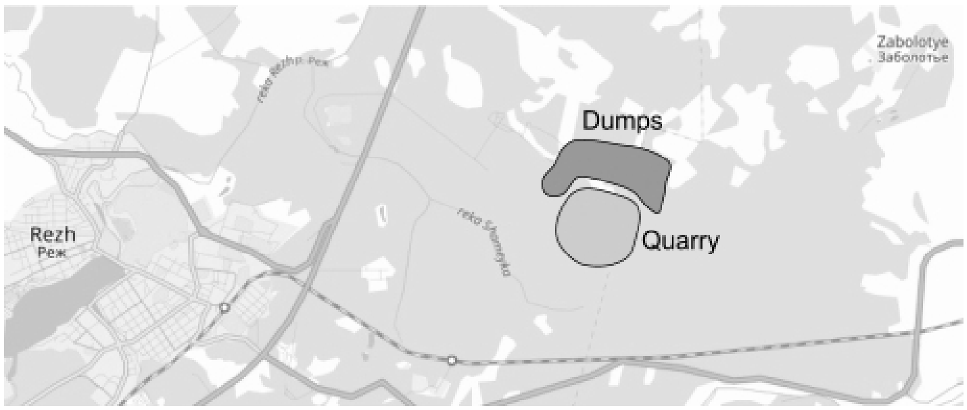

FIGURE 1. Location of Safyanovskoye copper pyrite deposit

(«SAS.Planet.Release.151111»)

In the course of research in 2018, 12 samples of herbaceous vegetation in the area of the studied object were selected. The layout of the sampling points is shown in Figure 2. The points are located both in the sanitary protection zone of the enterprise with the size of 500 m, and at a distance of 5-6 km from the quarry and waste dumps. Soil contamination is also monitored at these points, but these results are not discussed in this article.

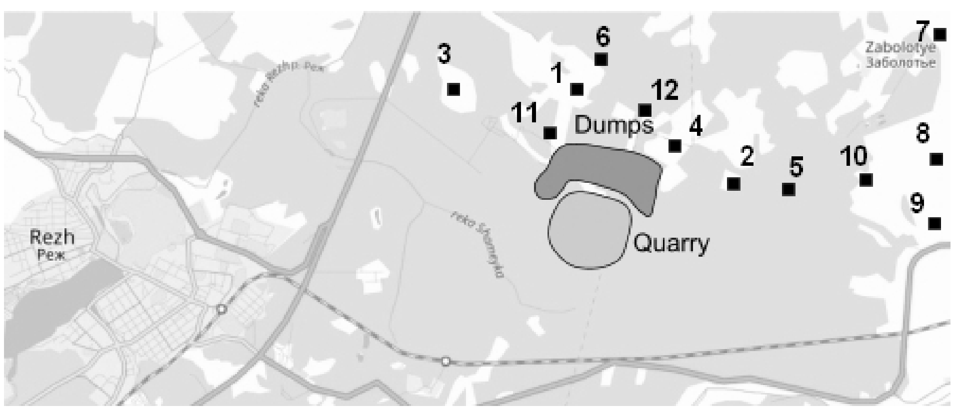

FIGURE 2. Soil sampling layout (squares)

(«SAS.Planet.Release.151111»)

Sampling and chemical analysis of vegetation samples were carried out in accordance with the current guidelines [17-19], as well as the requirements of the state standards of Russia. The results of the chemical analysis of vegetation samples are presented in Table 1. Since the selected samples are represented by samples of forage grasses and cereals, the obtained values of the content of pollutants were compared with the maximum allowed levels of heavy metals for coarse and succulent feeds (Maximum Allowed Levels - MAL according to [20]).

Chemical analysis of vegetation samples was carried out on the content of the main components that are specific to the area of the deposit: copper, zinc, nickel, lead, cadmium, arsenic. Copper and zinc are major ore metals (the content in the ore is from 0.5 to 2%). Arsenic, lead and cadmium are ore micro components (the content in the ore is less than 0.01%). Nickel can enter the territory from enterprises located in the town of Rezh at a distance of about 8 km from the deposit (metallurgical plant for the processing of nickel ores and nickel slag dumps).

TABLE 1. The content of chemical elements in samples of vegetation in 2018, mg / kg

# sample	Cu	Zn	Ni	Pb	Cd	As
1	8,80	16,40	4,4	3,10	0,09	0,34
2	13,20	30,20	13,30	1,90	0,28	0,55
3	14,40	26,90	4,90	3,90	0,10	0,48
4	14,30	74,60	8,00	2,10	0,16	0,35
5	10,70	24,50	6,90	0,57	0,07	0,44
6	10,20	19,60	4,10	3,20	0,12	0,51
7	12,00	22,30	6,60	0,88	0,13	0,43
8	12,30	23,30	5,20	1,30	0,70	0,41
9	9,60	19,80	4,50	6,70	0,19	0,51
10	11,20	23,20	5,90	1,40	0,08	0,51
11	12,70	121,40	5,70	4,20	0,28	0,47
12	13,60	83,10	9,50	3,10	0,25	0,50
MAL	30	50	3	5	0,3	0,5

During the monitoring and evaluation of the results obtained, an important aspect is the study of the influence of potential sources of pollution on the quality of vegetation in the study area and the separation of the zones of influence of these sources (such as quarry, overburden and host rock dumps, etc.). The mechanism of vegetation pollution at each specific monitoring point may be different, including a combination. Having determined the most likely mechanism of pollution, it is possible to identify a possible source of pollutants and assess how great its effect is on the vegetation of the study area.

According to monitoring data, the main mechanism of pollution is atmospheric transport of dust from quarries and dumps. Also, local pollution may be associated with the entry into the study area of sub-basement waters enriched with elements that are part of the ore, and the further transition of these elements to the ground part of grassy vegetation.

Cluster analysis of the results of chemical analysis of the selected vegetation samples was carried out to interpret the obtained data and to separate the observation points depending on their location in relation to the mining enterprise, as well as possible mechanisms of vegetation pollution.

In general, cluster analysis allows splitting objects not by a single parameter, but by a whole set of attributes. In addition, cluster analysis, in contrast to most mathematical-statistical methods, does not impose any restrictions on the type of objects under consideration, and allows us to consider a lot of initial data of almost arbitrary nature. At the same time, cluster analysis, like any other method, has certain disadvantages and limitations. In particular, the composition and number of clusters depends on the partitioning criteria chosen.

One of the problems of cluster algorithms is the estimation the distances between clusters or objects. When forming clusters of dissimilarity or distance between objects, the methods of joining or tree clustering method are used. It is also very important when conducting calculations to set the scale of measurements. Since different elements are characterized by significantly different background concentrations in vegetation, that is, different scales, the data to be analyzed must be standardized before the calculations are carried out to reduce to one scale.

When conducting cluster analysis, the data were normalized to MAL values (maximum allowable levels of vegetation pollution for coarse and succulent feeds) and standardized. Normalized to MALs and standardized data for analysis are presented in Table 2. Variables Var1-Var6 in Table 2 correspond to the following elements: Var1 - copper, Var2 - zinc, Var3 - nickel, Var4 - lead, Var5 - cadmium and Var6 - arsenic.

TABLE 2. Standardized data for cluster analysis

#	Var1	Var2	Var3	Var4	Var5	Var6
1	0,53	1,75	-0,18	-0,39	-0,88	-0,83
2	0,28	1,74	0,29	-0,69	-0,83	-0,80
3	0,57	1,77	-0,34	-0,44	-0,80	-0,76
4	-0,08	2,00	-0,30	-0,50	-0,57	-0,56
5	0,37	1,82	-0,03	-0,70	-0,75	-0,71
6	0,52	1,78	-0,29	-0,41	-0,83	-0,77
7	0,56	1,73	-0,05	-0,70	-0,79	-0,75
8	0,56	1,77	-0,22	-0,65	-0,72	-0,75
9	0,37	1,77	-0,33	-0,03	-0,92	-0,87
10	0,46	1,80	-0,13	-0,63	-0,78	-0,73
11	-0,24	2,03	-0,38	-0,42	-0,50	-0,49
12	-0,15	2,01	-0,27	-0,47	-0,56	-0,55

The cluster analysis was carried out in the standard version using the Statistica version 12 software package. The Unweighted pair-group average method was used as the merging rule. In this method, the distance between two different clusters is calculated as the average distance between all pairs of objects in them. The method is effective when objects actually form different groves, but it works equally well in cases of extended (chain type) clusters.

Euclidean distance was chosen as a measure of proximity. Euclidean distance is the most popular metric in cluster analysis. It is a geometric distance in multidimensional space. Geometrically, it best combines objects in spherical clusters. The measure of proximity, defined by the Euclidean distance, is a geometric distance in n-dimensional space and is calculated as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The dendrogram for 12 vegetation sampling points is shown in Figure 3. Variables C_1 - C_12 on the dendrogram correspond to sampling points 1 - 12.

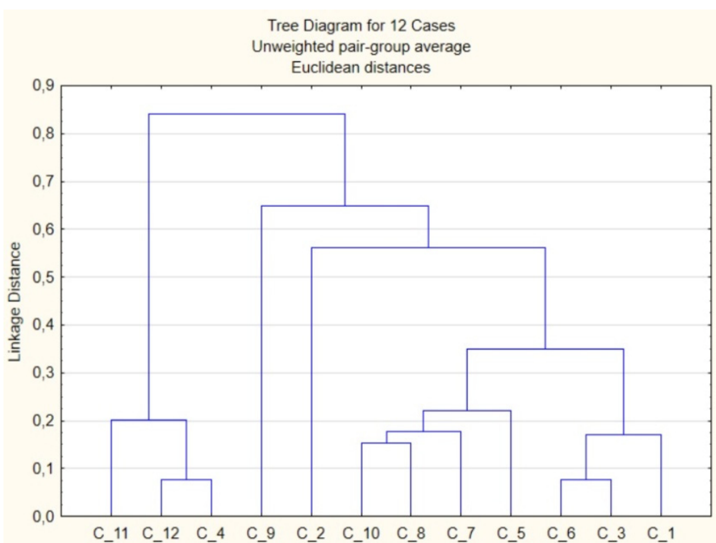

FIGURE 3. Dendrogram for 12 sampling points.

When studying the dendrogram for 12 sampling points, it can be assumed that the soil samples are forming three natural clusters. To verify this assumption, the initial data were divided by the K-means method into 3 clusters, and the significance of the difference between the groups was verified. The P-value for each variable was less than 0.05, which indicates the validity of this assumption.

The first cluster, obtained during the analysis, contains vegetation samples located in close proximity to the dumps and quarry (sanitary protection zone of the enterprise – up to 100 meters from the border of the quarry and dumps). That is, pollution at these points appears to be due to the entry of pollutants directly from the quarry site as a result of atmospheric transport of dust, as well as from the entry into the soil and further into the vegetation of sub-basement waters contaminated with soluble salts of ore metals and micro-components.

The second cluster contains samples of vegetation located at some distance from the quarry and dumps in the direction of the city of Rezh. Contamination of vegetation in the samples of this cluster can be associated with the flow of pollutants from industrial sources of the city of Rezh (the main sources are metallurgical aggregates of the nickel plant, emitting dust and gases from the smelting of nickel ore) during their atmospheric transfer to the studied territory.

The third cluster contains samples of vegetation that are significantly distant from the industrial area of a copper mine. The content of the studied metals in these points is rather low and, apparently, is related to the natural geochemical situation in the studied area. On the dendrogram we can also see two samples that are clearly not related to the resulting clusters. Given the low concentration of pollutants in the samples of these points and their location relative to the mine, these points can also be attributed to the third cluster.

CONCLUSION

The article presents the results of interpretation of monitoring data of agricultural vegetation in the area of a copper-pyrite deposit (August 2018). 12 vegetation samples were taken and chemical analysis was carried out to assess the contamination of vegetation with metals that are specific to the ore mined and for the area of the deposit. Then a cluster analysis of the chemical composition of samples of vegetation was carried out. During the analysis, all samples were divided into three clusters. This separation may be associated with different mechanisms for the entry of pollutants into the sampling sites, as well as with the location of the sampling sites relative to the industrial site of the enterprise. According to the results of the study, it can be concluded that cluster analysis is an effective tool for distinguishing the zones of influence of various sources of pollution of vegetation with a small number of measurements.

REFERENCES

1. Alexander M. Tarasyev, Julian Vasilev and Victoria F. Turygina. Quantitative analysis of raw materials mining of Sverdlovsk region in Russia, AIP Conf. Proc. 1738, 110009, 2016.
2. M.A. Medvedeva, A.N. Medvedev. About the Role of Environmental Monitoring in the Ensuring of Economic Safety of Mining Companies in Russia. 16th International Multidisciplinary Scientific GeoConference SGEM 2016, Conference Proceedings, Book 5, Vol. 1, pp. 19-26, 2016.
3. A. Medvedev, M. Medvedev, A. Sergeev. A Generalized Management Program by Environmental Monitoring Studies during Geological Prospecting. 16th International Multidisciplinary Scientific GeoConference SGEM 2016, Conference Proceedings, Book 1, Vol. 2, pp, 11-16, 2016.
4. Medvedev A.N., Sergeev A.P., and Melnichenka D.A., About Legislative Regulation of Local Environmental Monitoring in Russian Federation and Republic of Belarus, SGEM 2015, Vol. II, pp 591-597, 2015.
5. Emlyn E.F. Technogenesis of pyrite deposits of the Urals, Sverdlovsk: Publishing house of the Ural State University, 1991 (In Russian).
6. Saet Yu.E., Revich B.A., Yanin E.P. and others. Geochemistry of the environment, Moscow: Nedra, 1990 (In Russian).
7. Volkov S.N., Emlyn E.F., Ketsko O.G. The city of Rezh and its surroundings: nature, technology, people, Rezh-Ekaterinburg, 1992 (In Russian).
8. Tarasyev, A. M., Vasilev, J., & Turygina, V. F. Statistical analysis and forecasting of extraction and use of natural resources. AIP Conference Proceedings 2040, 050011, 2018.

9. Oleg I. Nikonov, Fedor P. Chernavin and Marina A. Medvedeva. The problems of classification: Method of committees, AIP Conf. Proc. 1738, 110004, 2016.
10. D.A. Tarasov, A.N. Medvedev, A.P. Sergeev, and A.G. Buevich. Review and Possible Development Direction of the Methods for Modeling of Soil Pollutants Spatial Distribution. AIP Conference Proceedings 1863, 050014 (2017);
11. M. Medvedev. On Improving the Efficiency of Environmental Monitoring: A Statistical Model of Snow Pollution by Different Sources. 16th International Multidisciplinary Scientific GeoConference SGEM 2016, Conference Proceedings, June 28 - July 6, 2016, Book 2, Vol. 1, pp. 775-780.
12. Manzhurov I.L., Berg, D.B., Antonov K. L. and Denisov D. I. Database of the regional waste cadastre of the Yamal-Nenets autonomous district. : CEUR Workshop Proceedings. 2109, pp. 43-48
13. M.A. Medvedeva and E. I. Komotskiy. About usage of data mining methods for fraud detection in the sphere of communal services, AIP Conf. Proc. 1738, 110011 (2016).
14. A.L. Gorbich, M.A. Medvedeva, and M.A. Medvedev. On application of vector optimization in the problem of formation of portfolio of counterparties. AIP Conference Proceedings 1789, 020005 (2016).
15. M. Medvedeva, S. Parusheva. Application of Information Technologies in Management of Greenhouse Gases and Wastes. 6th International Multidisciplinary Scientific GeoConference SGEM 2016, Conference Proceedings, June 28 - July 6, 2016, Book 2, Vol. 1, pp. 49-54.
16. Medvedev M.A. Information-Analytical System of Environmental Monitoring, SGEM 2015, Vol. I, pp. 685-690, 2015.
17. Program and methods of biogeocenological research. Edited by V.N.Sukachev and N.V.Dilisa, Moscow, 1966 (In Russian).
18. Field geobotany, Edited by E.M. Lavrenko and N.A. Korchagina, Vol. 1-4, Moscow-Leningrad, 1959-1972 (In Russian).
19. Short guide to geo-botanical research in connection with the field protective forestation and the creation of sustainable food supply in the European part of the USSR, Moscow, 1952 (In Russian).
20. Antonov B.I. Laboratory studies in veterinary medicine. Chemical and Toxicological studies. Guide. Moscow, «Agropromizdat», 1989 (In Russian).